IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF OKLAHOMA

STATE OF OKLAHOMA, **Plaintiff,** v. **TYSON FOODS, INC., et al.,** **Defendants.**	)	**Case No. 05-cv-329-GKF(PJC)**

DECLARATION OF JIM C. LOFTIS, Ph.D., P.E.

I, Jim C. Loftis, Ph.D., P.E. hereby declare as follows:

1. Since January of 1979, I have been a faculty employee of Colorado State University and am currently serving as professor in the Department of Civil and Environmental Engineering. My educational background includes a Bachelor of Science degree in Agricultural Engineering from Oklahoma State University and Master of Science and Doctor of Philosophy degrees in Agricultural Engineering from Colorado State University in 1976 and 1978, respectively. I am a registered Professional Engineer in the State of Colorado.

2. I have taught at least 20 different courses at Colorado State University, focusing on water and the environment in courses such as Environmental Statistics and Nonpoint Pollution. To serve the professional community, I have taught short courses in Water Quality Monitoring Network Design and Environmental Statistics in the U.S., New Zealand, and Australia.

EXHIBIT
"D"
tabbles

3. My faculty appointment as Colorado State University has involved a significant outreach and public education component through Cooperative Extension, and in 1990 my Extension activities in the area of agricultural impacts on water quality were recognized through an Outstanding Achievement Award from U.S. EPA Region VIII.

4. My research activities have focused in the area of environmental statistics including multivariate methods, design of water quality monitoring networks, and agricultural nonpoint source pollution. I have also conducted significant research in the areas of water resource system optimization and irrigation management. My research sponsors have included the National Park Service, US Environmental Protection Agency, the US Department of Agriculture, the US Geological Survey, the US National Science Foundation, and IBM Corporation. My recent research and consulting activities have included the following: serving as one of three experts on an external review panel for "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities Unified Guidance", 2004 Draft, U.S. Environmental Protection Agency; the design of water quality monitoring networks for the Big Thompson Watershed and the Upper Cache La Poudre River for northern Colorado water providers; statistical evaluation of reservoir and stream quality monitoring data for Colorado Front Range water providers; development of modeling and monitoring approaches for managing selenium contamination in the Gunnison River Basin; and evaluation of salinity sources and trends in the lower South Platte River basin.

5. In September of 2007 I was retained by counsel for the State of Oklahoma to provide statistical expertise and assistance with their analysis of water quality monitoring data for the Illinois River Watershed. I provided a peer review of Dr. Roger

Olsen's principal components analysis of the IRW data, including both his general methods and his specific application of those methods. I also reviewed Dr. Olsen's interpretation of the PCA results in light of (1) other lines of evidence that have been developed for this case (2) evidence developed outside of the work performed on this case and (3) my knowledge and experience of environmental processes, environmental statistics, water pollution in general and nonpoint source water pollution in particular. I am currently assisting Dr. Olsen and Dr. Richard Chappell with the preparation of a journal article documenting this work.

6. I have reviewed the scientific investigations and opinions of Dr. Roger Olsen contained in "Expert Report Of Roger L. Olsen, Ph.D." I have reviewed the opinions of Dr. Charles Cowan contained in his expert report "Rebuttal Report, Review of Principal Components Analysis of Data And Review of Inferences about Presence of Biomarkers in the Population of Animals from the Illinois River Watershed", November 26, 2008 and his deposition (Deposition of Charles Cowan, PhD, February 17-18, 2009). I have reviewed the opinions of Dr. Brian Murphy contained in his expert report "Expert Report of Brian L. Murphy, PhD", January 23, 2009 and his deposition (Deposition of Brian Murphy, PhD, March 25-26, 2009). I have reviewed the opinions of Dr. Glenn Johnson contained in his expert report "Rebuttal Report, Principal Components Analysis of Geochemical Data from the Illinois River Watershed Northwest Arkansas and Eastern Oklahoma", November 21, 2008 and his deposition (Deposition of Glenn Johnson, PhD, February 24-25, 2009).

7. Dr. Roger Olsen has utilized principal components analysis, PCA, to analyze and describe multivariate patterns in water quality across the Illinois River

Watershed, IRW. Multivariate methods, such as PCA, when applied to environmental studies, seek to condense the information from many different measured quantities into just a few indices that will reveal patterns, sources, and impacts more clearly than will the individual measurements. The PCA results have assisted Dr. Olsen in drawing conclusions regarding the impact of major pollutions sources, specifically land application of poultry waste and discharge of treated wastewater, on waters of the IRW. Dr. Olsen's approach is based on sound environmental science principles and well established, accepted statistical methodology. Based on this sound foundation, Dr. Olsen's opinions are reliable and have led to conclusions that are consistent with other major lines of evidence, including chemical mass balances, waste management practices, mass transport gradients, hydrologic modeling, and spatial analysis.

8. Principal components analysis or PCA, one of the simplest and most commonly applied multivariate statistical methods, has been widely applied in the environmental sciences, including water quality investigations such as the work done by Olsen in the IRW. The primary goals of principal components analysis are (1) to reduce the number of water quality variables (chemical and biological parameters) needed for explaining variability across the data set and (2) to identify and interpret variability and patterns. When there is considerable correlation among the variables (parameters) of interest, PCA produces just a few, ideally two or three, new variables that represent most of the variability of the entire set of different measurements (26 different variables for water in Dr. Olsen's case.) These new variables may be thought of as water quality indices which are weighted sums of the original variables or of appropriately transformed values thereof. The only difference between the indices that PCA produces and more

traditional water quality (or air quality, etc.) indices is that PCA computes the weights in a statistically optimal way so that all of the indices are independent, and the first principal component, PC1, represents the greatest possible fraction of the total variability; PC2 represents the next greatest possible fraction, and so on.

9. One of the common uses of PCA results in environmental studies is the identification of pollution or geochemical sources. To provide a brief snapshot of the most current applications of PCA in water quality investigations, I did a computer search of journal articles on May 28, 2009 using keywords "multivariate analysis of water quality data". I found the following four peer-reviewed articles published in scientific journals just in 2009. Many more have appeared in previous years, and Dr. Olsen based his specific approach to PCA directly on that large body of scientific literature. Articles published in 2009 include the following:

Kazi. T.G., M. B. Arain, M. K. Jamali, J. Jalbani, J. I. Afridi, R.A. Sarfraz, J. A. Baig, Abdul Q. Shah. 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study.Ecotoxicology and Environmental Safety 72: 301-309.

Iscen, Cansu F., A. Altin, B. Senoglu, and H. S. Yavuz. 2009. Evaluation of surface water quality characteristics by using multivariate statistical techniques: A case study of the Euphrates river basin, Turkey. Environmental Monitoring and Assessment 151:259-264.

Alkarkhi, A. F. M., J. Ismail, A. Ahmed, A. Easa. 2009. Analysis of heavy metal concentrations in sediments of selected estuaries of Malaysia- a statistical assessment. Environmental Monitoring and Assessment 153:179-185.

Qi Zhang, A. Li, G. Zeng, J. Li, Y. Fang, Q. Yuan, Y. Wang, F. Ye. 2009. Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. Environmental Monitoring and Assessment 152:123-131.

10. Like Dr. Olsen, these investigators use principal components analysis based on the matrix (table) of correlations among the multiple water quality variables. One study used a varimax rotation of the principal components. Dr. Olsen also examined rotated factors in his analysis. These studies used the PCA loadings of the individual variables (parameters) to relate individual principal components (e.g., PC1 or PC2) to specific types of source waters or pollution sources. The approach is illustrated by the following quote from Zhang et al. (2009), above.

> "For the data set representing the MP sites, among total four significant PCs, PC1, explaining about 25.51% of the total variance, has strong positive loading on Cd and moderate positive loadings on Cu, As and Pb. These factors represent the contribution of the industrial from heavy industry. PC2, explaining about 17.29% of the total variance, has strong positive loading on, COD and moderate loadings on DO and ammonia nitrogen, which represent pollution from domestic wastewater of Xiangtan City and Loudi City. PC3, explaining about 15.80% of the total variance, has strong positive loadings on nitrate nitrogen and total phosphorus, moderate loading on BOD5. In these areas, farmers use the fertilizer, which represents point and non-point source pollution from orchard and agriculture areas. PC4 (13.15%) has moderate positive loadings on PH and Cr. The factor represents the pollution of electroplating wastewater."

11. This is the very same approach that Dr. Olsen has used to associate his PC1 and PC2 for the surface water data sets to land application of poultry waste (PC1) and to treated wastewater discharges (PC2). Also like Dr. Olsen, these and similar studies in the literature linked spatial analysis with principal component analysis to evaluate water quality patterns and identify sources.

12. The opinions of Defendants' experts Dr. Murphy and Dr. Johnson are at odds with the current scientific literature on this type of PCA analysis, recommending instead a different approach that is based on ratios rather than actual concentrations of the various constituents. Their approach makes it difficult if not impossible to ascertain the

degree of contamination that is present at a given location. Their approach derives from their experience in the areas of impacts of synthetic chemicals, like PCBs. Neither Dr. Johnson nor Dr. Murphy has significant experience in the area of nonpoint source pollution from agriculture and naturally occurring chemicals, like phosphorus. Defendants' other statistical expert, Dr. Cowan, has no experience in environmental data analysis at all.

13. Dr. Olsen's approach follows a well trodden path established for many years in the scientific literature. His is the standard approach to multivariate analysis of water quality data where naturally-occurring constituents are the concern, and in no respect is a new or novel approach or methodology. His approach has been tested, proven, and validated time and time again in the scientific literature.

14. Dr. Olsen's PCA results are readily reproducible. As part of my peer review of the PCA study, I independently computed principal components for Dr. Olsen's data set SW_15, which consists of 267 cases, each with 26 water quality variables. The cases span a wide range location types, flow conditions (low flow and high flow), and degree of water quality impacts—from locations that are not immediately downstream of known pollution sources (and including reference areas) to locations that are immediately downstream of poultry waste land application or wastewater treatment plant discharges. The water quality variables include an appropriate collection of both chemical and bacterial measurements. For a data set of this type, I expected and found both significant correlation among variables and large variability of most water quality variables across cases. Both of these characteristics make these data well suited for PCA.

15. Like Dr. Olsen's, my initial evaluation of the data included the calculation of basic summary statistics, both on the untransformed and log transformed data. The summary statistics included the mean, median, range, coefficient of variation, and skewness coefficient. Since the untransformed data tended to be significantly right-skewed for all variables except calcium, the log transformation is appropriate. The log transform results in much less skewness and, therefore, results in distributions closer to the normal distribution. While a normal distribution is not assumed or required in the PCA as applied here, data that are closer to "normal" are desirable in that variance calculations will be less susceptible to extreme influence of a few outlying values. It is also desirable to transform all of the variables in the same way (i.e. log transform) when that is reasonable as it is here.

16. The use of log transformed data in the analysis is absolutely critical for bacteria and other highly skewed variables. Otherwise a few extreme values will inappropriately dominate the analysis in that the first (few) principal components will be derived primarily to separate out samples with extreme values from the rest. (See Pierre Legendre and Louis Legendre, 1998. *Numerical Ecology,* 2nd English Edition, Developments in Environmental Modeling, 20. Elsevier, p. 411.) Many PCA studies of water quality that do not include bacterial variables do not log transform the data. However, Dr. Olsen chose to use a log transform for all variables for the reasons I mentioned above. As R. O. Gilbert notes in his widely used text, *Statistical Methods for Environmental Pollution Monitoring* (John Wiley and Sons, 1987) "The lognormal distribution is the most commonly used probability density model for environmental data".

17. My check of Olsen's PCA included computing principal components equations and loadings for all 26 variables, saving only the first two principal components because those two accounted for a large fraction of the total variance. My calculations were performed using the Minitab statistical software package, while Dr. Olsen used SYSTAT. The principal component equations and loadings that I obtained were identical to those obtained by Dr. Olsen.

18. I also performed a sensitivity analysis of the PCA results. There were multiple phosphorus (P) variables (total P, total dissolved P, and soluble reactive P) and multiple bacteriological variables (total coliforms, fecal coliforms, *E.Coli*, Enterococcus) included among the 26 which could contain some redundant information. While PCA effectively accounts for correlation between variables, I was interested in whether very similar results would be obtained if the analysis was limited to one phosphorus variable (total P) and one bacteria variable (total coliforms). The analysis with one phosphorus and one bacteria variable produced essentially the same principal components in terms of the order of importance of each variable in the first two principal components.

19. Dr. Olsen performed a much larger number of sensitivity runs to test the potential effect of choices on which variables to include and how to manage missing values and nondetects. Through these sensitivity runs, Dr. Olsen has demonstrated that the PCA results are robust against (meaning relatively unaffected by) these decisions.

20. Surprisingly, defendants criticizes the fact that that Dr. Olsen performed a very large number of runs and that "Dr. Olsen always sees a poultry litter "signature" that no other scientist has ever seen." (Defendants' motion p. 8). Defendants claim that Dr. Olsen performed many runs in order to be able to hand pick those that showed a

poultry signature. However, Dr. Olsen's expert report is clear that these multiple runs were performed to determine sensitivity of the PCA results to how the data were handled and PCA performed. As I discuss below, these very issues were raised by Dr. Cowan, who clearly did not consider the sensitivity runs in his expert report. Thus, there is no basis for Defendants' criticism of either the large number of runs or of the data handling issues that these runs were used to address.

21. The assertion that no other scientist has seen the poultry signature is extremely misleading, implying that many other scientists have followed this same line of investigation and found nothing. This is simply not true. In fact, many, many scientists have observed poultry impacts in the IRW from whatever perspective they looked or in which they had expertise. While Dr. Olsen may be the first to use principal components analysis in the IRW, his conclusions regarding the significance of poultry waste impacts are not different from the conclusions of other scientists, including those from state and federal agencies, who have studied IRW water quality. His conclusions are also consistent with other source-identification methods used in this case, such as mass balance, concentration gradients, and watershed modeling.

22. Dr. Olsen's most important findings from PCA are very simple, logical, and consistent with other methods of analysis. There are two major findings. First, Dr. Olsen found that a large fraction of the variability in water quality across the IRW (based on 26 variables measured at many locations from edge of field to Lake Tenkiller) could be explained by two indices, PC1 and PC2, that could logically be associated with poultry waste application to land and municipal wastewater discharges, respectively. The association of the PCs to these particular sources is based primarily on a comparison of

the weights (also called loadings) for each water quality variable (constituent) in each PC with the importance of the same constituent which was also found in the waste, as described above by Zhang et al. (2009). More specifically, the constituents with the largest weights in PC1 match up well with the constituents of greatest importance in both poultry litter and, more importantly, in edge of field samples from fields where poultry waste has been applied. Similarly, the weights for individual constituents in PC2 match up well with the composition of treated wastewater.

23. Defendants' experts have not been able to refute or even really challenge this association between the observed loadings of PC1 and PC2 and the composition of the corresponding waste sources, which is one of the key bases of Dr. Olsen's conclusions. Instead, Dr. Johnson and Dr. Murphy have relied upon claims that loadings do not equate to sources. (Dr. Cowan does not discuss the issue directly.) Obviously, this claim is true. The association or relationship does not imply equivalence, nor does Dr. Olsen suggest differently. It is rather the relationship between PC loadings and unique source compositions of both poultry waste and treated wastewater that is used by Dr. Olsen. The same relationships are used by the authors cited earlier and many other scientists in the literature.

24. Dr. Olsen's second major finding is that PC1 scores are spatially distributed in a pattern that is consistent with increasing poultry waste impact in water bodies that are closest to areas of poultry waste application and decreasing impact as one moves downstream, or in areas with little poultry waste application upstream. Defendants' expert, Dr. Johnson, spends a large amount of his expert report disputing the results of individual points in Olsen's spatial analysis. Dr. Johnson's deposition clearly

shows that his own interpretation of the spatial analysis is not based on an understanding of where poultry litter has actually been applied and is, therefore, inconsistent with the simplest form of fate and transport analysis, i.e. pollutants move directly downstream from sources.

25. Dr. Johnson's focus on whether a particular dot on Dr. Olsen's maps should be red or green is, intentionally or not, at odds with the purpose and utility of spatial analysis of PCA scores. The goal of spatial analysis is to reveal overall patterns of water quality in a multivariate context that would be less obvious if the variables were considered one at a time. Dr. Olsen's analysis of PC1 scores shows simply that higher values of PC1, which suggest greater impact of poultry waste disposal, generally occur closer to the areas in which poultry waste has been land applied. These scores decrease in a logical fashion as one moves downstream toward Lake Tenkiller and are lowest in reference samples. The numerical value that one chooses as a cutoff value of PC1 scores for grouping "predominantly poultry impacted" sites is not critical, in and of itself, though it is easy to criticize. What is important is the general patterns in the scores and what they reveal about the general patterns in water quality. Dr. Olsen's spatial analysis is consistent with and reinforces his conclusions regarding the association between PC1 and poultry waste application. His spatial analysis is also consistent with the association between treated wastewater discharges and PC2 that is apparent in the PC2 loadings values. Thus the PCA results are internally consistent. The number and locations of green or red dots on Dr. Olsen's maps are reflective of the scope and extent of the impact of poultry waste land application in the IRW.

26. Even more important, however, is that Dr. Olsen's conclusions from the PC analysis are also consistent with the other methods of analysis that he relied upon in developing his overall conclusions regarding the impact of poultry waste on IRW water quality. These methods include

1) Mass balances, including identification and quantification of the major waste types discharged into the IRW, including poultry, cattle and wastewater treatment plant discharges. Mass balance analysis reveals that most of the dozens of minor sources of contamination are insignificant and need not be considered.
2) Watershed modeling, using a basin-wide model to determine the amounts of phosphorus from various sources.
3) Fate and transport evaluations, focusing on the amount and locations of waste application; geology and hydrology of the IRW; concentration gradients of individual constituents, especially phosphorus, along mass transport pathways from poultry waste application areas to Lake Tenkiller; and the evaluation of chemicals in Lake Tenkiller sediments and reference locations.
4) Results of a poultry-specific bacterial biomarker analysis.

27. Defendants' expert, Dr. Johnson, is particularly critical of Dr. Olsen's failure to find a cattle signature that is as obvious as the poultry signature. While it would be very nice if PC3 or PC4 were obviously associated with cattle, that is not the case. Neither is it surprising for two reasons. First cattle in the IRW are fed primarily on grass, which is a natural food source much different from that fed to poultry. Thus cattle manure is much different in composition from poultry waste, with smaller phosphorus content and smaller leachable contaminant mass. Second the grass that cattle eat in the

IRW is mostly grown on fields where poultry waste is applied. The poultry waste supplies nutrients and other chemicals to the grass that reappear in cattle manure. Thus, it is not surprising that the impacts of cattle on water quality, as revealed by PCA, are not as distinct as those of poultry, or, by similar reasoning, as those of wastewater discharges.

28. Defendants' experts, Dr. Cowan and Dr. Johnson, assert that Dr. Olsen's incorrect computation of PC scores (incorrect because raw data were used instead of logs to compute scores) results in greatly different conclusions compared to those obtained when the scores are computed correctly, using the logs. As I stated above, the most important conclusions from Dr. Olsen's PCA analysis, or any PCA analysis for that matter, are not those revolving around the numerical values of the scores. The important conclusions are rather those concerning the patterns of water quality that are more clearly discerned and displayed by a multivariate analysis than by considering each variable separately. The inadvertent mistake, now corrected by errata, did not affect the general and important conclusions at all.

29. The first major finding that I mentioned above is related only to loadings, not scores, and is thus totally unaffected by the computational error. The second major finding regarding the spatial distribution of PC1 scores is unaffected as well. The highest scores occur closest to areas of poultry application and decrease as one moves downstream and away from areas of application. There are two reasons that the overall conclusions are not affected. The first is that the log transform does not affect the order of the observations. The largest value has the largest log and so on. The second is that all observations are further transformed into normal scores before insertion into the equations for the principal components. This normalization places all of the observations

for all variables into a common, dimensionless reference frame in which the range of values is roughly -4 to +4, regardless of the measurement. This is the most common approach for computing both principal components and scores and allows inclusion of different types of measurements with varying units and greatly varying magnitudes.

30. The grouping of points from various types of source waters is unaffected by the computational error as well. In both cases, in a scores plot of PC2 vs. PC1 for surface waters, the edge of field samples are grouped together, the wastewater impacted samples are grouped together, and the reference (unimpacted) samples are grouped together. (See original and corrected figures 6.11-18 c and d for Dr. Olsen's run SW3.) In these figures, ambient waters plot somewhere in the middle, indicating varying contributions from the various types of source waters. This type of plot is commonly used in PCA studies of water quality to aid in identification of sources and to identify the relative contributions of particular types of source waters.

31. Defendants' expert, Dr. Cowan is highly critical of several data-handling aspects of Dr. Olsen's specific application of PCA. These include Dr. Olsen's method of handling missing data, nondetects, and laboratory duplicate observations. As described below, all three aspects are very common in water quality data sets, and they deserve careful consideration in order to derive valid conclusions from water quality data. Dr. Olsen has a very well thought out protocol for dealing with all three issues that is clearly stated in his report and is consistent with accepted practice.

32. Missing data are common in water quality data sets, primarily because not all constituents are measured in every sample. This is usually intentional. The analysis frequency for each variable is determined based on its importance, its variability and

sometimes on the cost of analysis. Dr. Olsen carefully manages the impact of missing data by strictly adhering to stated limits on the maximum number of missing values that will be permitted for a given water sample to be included in the analysis. For samples that meet these criteria, Dr. Olsen computes PC scores for samples with missing data by using a normal score value of Z=0, corresponding to the mean of the missing constituent. Defendants statement that "This substitution actually *increases* the overall potential variability in the dataset, and distorts the resulting principal component." (Defendants' Motion p. 21 and Ex. 3 at 22-25) -is false. Substitution of the mean (Z= 0.0) makes a zero contribution to the principal component score, to the variance of the principal component, to the variance of the individual constituent, and to the variance of the overall data set.

33. Non-detects are very common in water quality data because studies typically include variables that occur at very low concentration in at least some samples. Dr. Cowan clearly does not understand what non-detects are and how they are commonly handled in environmental studies. Dr. Cowan's report implies that a non-detect demonstrates the absence of the constituent being detected. (Defendants' Motion Ex. 3, p.26) In fact, the constituents of concern in this PCA are all ubiquitous, naturally occurring substances, for which the complete absence in any field sample is virtually impossible. A recorded non-detect simply means that the concentration measured by the laboratory is below the method detection limit for that analysis. The method detection limit is a statistically determined concentration below which the signal to noise ratio for the measurement is unacceptably low. Below this level, the laboratory simply records a

"less than X", where X is the numerical value of the detection limit. This does not mean zero.

34. The standard and by far the most common approach to dealing with non-detects in principal components analysis is to substitute half of the detection limit, which is what Dr. Olsen does. According to the Defendants' motion (p. 23) this approach "creates data where none were measured". That is, of course, false, since there is an actual laboratory measurement behind every recorded non-detect. The measurement is simply below the method detection limit established by the laboratory and is, therefore, recorded as a "less-than."

35. Dr. Olsen limits the effects of non-detects in his analysis by not including any variable for which there is a high percentage of non-detects. There is no point in including variables that are mostly non-detects anyway since PCA is all about analyzing variability. Observations that are mostly non-detects do not contribute very much to the total variability because they are all statistically analyzed as the same value, usually half the detection limit.

36. The following quote from defendant's motion (p. 23) is based on Dr. Cowan's report (Defendants' Motion Ex. 3, p.32):

> "Second, the fact that Dr. Olsen used different detection limits for the same constituents further distorts the analysis. *Id.* at 26, 32-33. While small in absolute numbers (*e.g.*, 0.01 vs. 0.001), these differences are magnified by Dr.Olsen's use of logarithmic values in his PCA runs. When converted to a log scale, 0.01 becomes -2, 0.001 becomes -3, and so forth, which can have a substantial effect on the outcomes."

This statement demonstrates Dr. Cowan's fundamental lack of understanding of what non-detects are, the applicability of logarithms, and how to work with different units of measurement—all elementary concepts for analyzing environmental data. First, Dr.

Olsen does not establish the detection limits. They are established by the laboratory, and can and commonly do vary over time within the same laboratory for the same constituent. Second, the use of logarithms does not distort anything. The values 0.01 and 0.001 are not close together. They differ by a factor of ten; thus they differ in log units by 1.0. To see this more clearly, detection limits of 0.01 or 0.001 are most likely in units of milligrams per liter or parts per million. One could just as easily use units of micrograms per liter or parts per billion, in which the same exact detection limits will be 10 and 1 respectively, and the corresponding logs will be 1.0 and 0.0. The outcome of any statistical analysis must be the same, regardless of the system units used, and in any system of units, one log unit will correspond to a factor of ten. In short, the effect of multiple detection limits is not exaggerated by taking logarithms. In all respects, Dr. Olsen has handled non-detects appropriately.

37. Dr. Cowan criticizes Dr. Olsen for averaging laboratory duplicates for the same sample, rather than treating them statistically as separate observations. This unusual and surprising complaint from Dr. Cowan underscores his lack of experience in dealing with environmental data that include laboratory duplicates. Laboratory duplicates are intentionally generated and are in fact absolutely required for quality control purposes. The standard approach with environmental data sets is to average the duplicates. One might also logically pick one value at random. To include all of them as separate data points is a serious mistake since that will result in overweighting the sample for which duplicate measurements are made. Clearly Dr. Olsen is correct and is following accepted practice in averaging laboratory duplicates.

38. In addition to criticizing data handling, Dr. Cowan also criticizes Dr. Olsen for merging data from both USGS and CDM in his analysis. The data are said to be incompatible because one obtains different results if one analyzes each data set separately. In actual fact Dr. Olsen's data set intentionally includes several different types of data from different flow regimes and different points along the flow path. The CDM and USGS data are based on very similar analytical methods and are comparable in that sense. However, the USGS data represent the larger rivers with larger flows. The CDM stream monitoring program was intentionally designed to supplement and fill gaps in the USGS program by representing the smaller streams farther up the watershed. Also included in the PCA are edge-of-field samples and lake samples, which represent the extreme ends of the flow path. For some runs, ground water data are included as well. Obviously if one analyzes any of these individual data sets separately, one will get different results than if they are analyzed together. But analyzing them together gives an overview or big picture of how water quality changes along the flow path in a multivariate sense. The inclusion of both USGS and CDM data within the same PCA is correct and necessary to provide the complete picture of water quality patterns that Dr. Olsen seeks.

39. To summarize, Dr. Olsen's approach to PCA is based on numerous journal articles that have shown this methodology to be a technically sound and effective means of re-expressing the information contained in a large number of variables in a much smaller number of indices called principal components. This methodology has been widely used and validated in the literature for identifying and characterizing the impacts of pollutant sources. I have independently examined Dr. Olsen's specific approach and

reproduced his calculation of principal component equations and loadings for his data set SW_15, using a different software package from the one that Dr. Olsen used. Through preliminary data analysis, I verified that Dr. Olsen's use of the log transform for this data set is appropriate. Using extensive sensitivity analysis, Dr. Olsen thoroughly tested and effectively demonstrated the robustness of the analyses to the manner in which the data were handled and the PCA was performed. Dr. Olsen has clearly performed his PCA correctly, using currently accepted and published methodology. Therefore, his specific application was correct also. His overall conclusions regarding the impacts of poultry waste application in the IRW are based both on the PCA results and other standard methods of pollutant transport and fate analysis that consider water quality variables one at a time.

I declare under penalty of perjury, under the laws of the United States of America, that the foregoing is true and correct.

Executed on the 29th day of May, 2009.

Jim C Loftis

Jim C. Loftis, Ph.D

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF OKLAHOMA

STATE OF OKLAHOMA,	)	
	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 05-cv-329-GKF(PJC)**
	)	
TYSON FOODS, INC., et al.,	)	
	)	
Defendants.	)	

DECLARATION OF RICK W. CHAPPELL, Ph.D.

I, Rick W. Chappell, Ph.D., hereby declare as follows:

1. Since June 2005, I have been an independent contractor and owner of Environmental Science Solutions LLC, an environmental consulting company. My educational background includes a Bachelor of Science degree with special distinction in Geology from the University of Southern Colorado (1980) and a Doctor of Philosophy degree in Geochemistry from the Colorado School of Mines (1989).

2. Since 1997, I have been an adjunct professor and instructor in the Civil Engineering Department at the University of Colorado at Denver and Health Sciences Center, where I have taught courses in water quality laboratory and environmental engineering. These courses included sections on statistical analyses.

3. From 1985 to 2005, I was an environmental scientist at Camp Dresser & McKee Inc. ("CDM") in Denver, Colorado, where I was a senior geochemist and environmental statistics leader. As a geochemist, I have extensive experience performing environmental fate and transport studies and evaluating the sources of contamination in

EXHIBIT
"E"
tabbies

the environment. As CDM's environmental statistics leader, I reviewed statistical analyses conducted by CDM staff, wrote technical guidance documents for statistical analyses, taught training courses on environmental statistical analyses, and developed computer programs to conduct environmental statistical analyses. Training courses I have taught include: a workshop for the U.S. Agency for International Development ("Environmental Data Analysis"); a symposium for the New England Water and Environment Association ("Geostatistical Applications for Environmental Investigations"), two training courses for the Michigan Department of Natural Resources ("Environmental and Statistical Data Analysis"), and three training courses for CDM ("Environmental Statistics" and "Fundamental Geostatistics for Environmental Investigations"). Recently I developed a new training course for CDM ("Environmental Statistics with Application to Censored Data") that I will be teaching at four locations across the United States. Although now the owner of my own company, I have continued my role as CDM's environmental statistics leader. I have over 20 years of experience in planning, developing, and implementing statistical methods for environmental investigations. My experience includes the statistical method called principal components analysis ("PCA").

4. In November 2004, CDM was retained by the Oklahoma Attorney General to perform a scientific investigation concerning environmental contamination found in the Illinois River Watershed ("IRW"). I have been involved in this investigation, first as a CDM employee, and then as a subcontractor to CDM, since 2005. My primary role on the IRW project has been to support CDM staff and to conduct statistical analyses. In this capacity, I have assisted Dr. Roger L. Olsen, Ph.D. in conducting a PCA, one of

several scientific methods applied by Dr. Olsen and other State of Oklahoma (State) experts in evaluating contamination in the IRW.

5. Dr. Olsen's results and opinions regarding the IRW investigation are contained in his expert report ("Expert Report of Roger L. Olsen, Ph.D., May 14, 2008") and deposition transcript ("Deposition of Roger L. Olsen, Ph.D., September 10-11, 2008"). I have thoroughly reviewed both of these documents. I have also reviewed the opinions of Defendants' experts contained in rebuttals of Dr. Olsen's expert report and in deposition transcripts. I have reviewed the opinions of Dr. Charles Cowan contained in his rebuttal report ("Rebuttal Report, Review of Principal Components Analysis of Data and Review of Inferences about Presence of Biomarkers in the Population of Animals from the Illinois River Watershed, November 26, 2008") and deposition transcript ("Deposition of Charles Cowan, PhD, February 17-18, 2009"). I have reviewed the opinions of Dr. Brian Murphy contained in his rebuttal report ("Expert Report of Brian Murphy, PhD, January 23, 2009") and deposition transcript ("Deposition of Brian Murphy, PhD, March 25-26, 2009"). And I have reviewed the opinions of Dr. Glenn Johnson contained in his rebuttal report ("Rebuttal Report, Principal Components Analysis of Geochemical Data from the Illinois River Watershed, Northwest Arkansas and Eastern Oklahoma, November 21, 2008") and deposition transcript ("Deposition of Glenn Johnson, PhD, February 24-25, 2009").

6. Section 6 of Dr. Olsen's expert report ("Evaluation of the Sources of Contamination in the IRW") presents the results of multiple lines of scientific evidence obtained by the States' experts. These lines of evidence are studies that represent the individual components of an environmental fate and transport analysis. They include the

following: a study of the geology and hydrogeology of the IRW in relation to the fate and transport of potential contaminants and sources of contamination; a study of the chemical and bacterial composition of potential contaminant sources; a study of the potential sources and mass balance of phosphorus, bacteria, and other contaminants in the IRW; a study of the chemical and bacterial contaminants observed in each environmental component of the IRW; a study of the nature and extent of contamination observed in surface water, groundwater, and sediments, including core sediments from Lake Tenkiller; a study of the fate and transport of poultry-related contaminants; a study of poultry waste biomarkers; a study of the amounts of wastes generated by all potential major sources in the IRW and their methods of disposal; a study of contaminant concentration gradients, or the trends in contaminant concentrations going from upstream to downstream locations in the IRW; a watershed and reservoir modeling (computer simulation) study for phosphorus in the IRW; and a study of the chemical and bacterial signatures using PCA. Together, these studies constitute a comprehensive environmental fate and transport analysis.

7. PCA is a type of multivariate statistical analysis commonly applied in environmental science. It is conducted in order to assist the environmental scientist with the identification of correlation patterns amongst multiple samples and multiple variables (in environmental science, the term "variable" is also referred to as "constituent" or "parameter" and represents chemical or bacterial concentrations measured in samples). These correlation patterns are shaped by the underlying environmental factors comprising the particular environment under study. When the factors shaping them are due to sources of contamination, the patterns produced by PCA reflect these sources of

contamination. Many textbooks have been written on multivariate statistical analysis. One text that I find useful, particularly for my students, concerns the direct application of multivariate statistics in environmental science: "Multivariate Statistics for the Environmental Sciences" by Peter Shaw (2003). Chapter 6 of Dr. Shaw's text is devoted to PCA and its application in environmental science. Dr. Shaw notes that PCA is "…the ordination technique of first choice for the majority of environmental scientists." (In environmental science, the term "ordination" means "finding order within data", and in the context of PCA, it is equivalent to "correlation pattern").

8. The number of scientific papers written on PCA and its application in environmental science is voluminous. In many cases the investigators have applied PCA to environments similar to the IRW. In Tables 6.11-1a and 6.11-1b of his expert report, Dr. Olsen lists 32 scientific papers where PCA was applied to environmental studies. He also provides a direct comparison to the IRW of four watersheds where PCA was similarly applied (Table 6.11-2). Other applications of PCA to watersheds include: the Lake Lanier watershed ("Multivariate statistical characterization of water quality in Lake Lanier, Georgia, USA, J. Environ. Qual. 34: 1980–1991, Zeng and Rasmussen, 2005") and the Richibucto River watershed ("Multivariate analysis of water quality in the Richibucto drainage basin, New Brunswick, Canada, J. American Water Resources Assoc. 40(3): 691-703, St-Hilaire and others, 2007"). These watersheds are comparable to the IRW in terms of size and number of years of monitoring. In all of these studies, the investigators used PCA as a means of characterizing sources of contamination. This is important because Defendant's motion to exclude Dr. Olsen's testimony (the motion)

states that the application of PCA to characterize sources of contamination in watersheds like the IRW is novel and untested. In fact, it is not.

9. Section 6.11 of Dr. Olsen's expert report describes the methodology he followed in conducting PCA. The methodology includes the following steps: extraction of pertinent data from the project database; reduction of extracted data into PCA datasets; exploratory data analysis and data transformation; and running of PCA. This methodology is followed by all environmental scientists conducting PCA. In the IRW investigation, the project database is a Microsoft ACCESS computer file named "IllinoisMaster.mdb". This master database file is the repository for all analytical data collected from every individual study conducted over the course of the IRW investigation. As explained later, data were collected for many purposes, so not all data contained in the master database file is applicable to PCA only. The first step in the methodology is to extract the pertinent data. This is accomplished by constructing database queries, which are a set of instructions that are followed to extract the pertinent data. The queries used by Dr. Olsen are retained in the master database file so that they can be examined and followed by another environmental scientist. The master database file (containing these queries) was provided by Dr. Olsen in his considered material. This is an important point because Defendant's expert Dr. Cowan argues that he was unable to reproduce Dr. Olsen's surface water PCA dataset. I have performed a thorough evaluation of this issue and determined that Dr. Cowan ignored the queries used by Dr. Olsen. My evaluation of this and other issues regarding Dr. Cowan's reproduction of the surface water PCA dataset are contained in my previous declaration (Exhibit E to Docket

2072), which is also provided in Attachment A. I will discuss many other problems and mistakes of Dr. Cowan in the following paragraphs.

10. The IRW investigation, like any large environmental investigation of its kind, resulted in minor errors that were not identified by Dr. Olsen until after he submitted his expert report. These errors were corrected by Dr. Olsen as detailed in three "Errata" documents. I have thoroughly examined these documents. Most of the errors were minor typographical errors. One error was a programming error that occurred near the end of his PCA methodology. All identified errors were corrected by Dr. Olsen. All of the errors were minor and therefore had no impact on his opinions. Correction of the programming error actually improved the PCA results, allowing Dr. Olsen to make an even more definitive characterization of contaminant source impacts in the IRW.

11. With regard to Defendant's expert Dr. Cowan, I have concluded from his rebuttal report and deposition transcript that he is not an environmental scientist. He therefore lacks the necessary experience with environmental investigations to provide an adequate review of Dr. Olsen's work. My conclusions concerning Dr. Cowan's lack of experience are further detailed and supported in my previous declaration (Attachment A). This is important because the Defendants continue to use Dr. Cowan's opinions to criticize Dr. Olsen's PCA, even though Dr. Cowan is not qualified.

12. As discussed previously, Dr. Olsen designed the IRW investigation to support multiple scientific studies, and these studies constituted the components of an overall environmental fate and transport analysis. Each study had different objectives and requirements. The objectives and requirements of an environmental investigation are established and evaluated according to a standard process referred to as the data quality

objectives (DQO) process. The DQO process is followed to ensure that the necessary types, quantity, and quality of data are obtained to support the requirements of the particular study. The U.S. Environmental Protection Agency ("EPA") provides guidance documents for the DQO process ("EPA Quality Manual for Environmental Programs, EPA Order 5360 A1"; "Guidance for the Data Quality Objectives Processs, EPA QA/G-4"). I am thoroughly familiar with these guidance documents. Dr. Olsen provides the DQO process followed in the IRW investigation in Section 3 of his expert report. The IRW DQO process is consistent with EPA guidance. One of the important elements of the DQO process is incorporation of a phased approach to data collection. In a phased approach, data are evaluated at various points along the way to ensure that the data are meeting (or are continuing to meet) the requirements of each individual study. If not, data gaps are identified and the investigation is modified accordingly (such as by collecting additional samples or analyzing samples using a different laboratory method). Dr. Olsen and other States' experts followed this phased approach. They conducted analyses and evaluations at various points during the course of the IRW investigation. Then, as necessary, data collection and data analysis programs were modified to ensure that data of sufficient type, quantity, and quality were obtained. Again, each study generally had different objectives and requirements. Therefore, the data collected (the types and quality of chemical analyses) were not necessarily the same across each study, nor were they intended to be the same. This is important because Defendants' experts (particularly Dr. Cowan) have used this fact as a basis for criticism of the PCA. They argue that thousands of samples were collected, for more that 100 different variables (constituents), but that not every single sample was analyzed for every single variable.

They argue that because not every single sample and not every single variable was included in the PCA that the PCA must therefore be invalid. First, such arguments underscore the ignorance of the Defendants' experts with regard to the DQO process. It was simply not necessary to analyze each sample for every possible variable (constituent). Nor was it necessary that every sample collected be included in the PCA. This is because many samples were collected and analyzed for the purpose of other studies, and not for the PCA. Dr. Olsen obtained comprehensive and representative data for the PCA. He followed the DQO process in doing so. Second, the argument that over 100 different variables (constituents) were analyzed is a mischaracterization and underscores the ignorance of Defendants' experts regarding even a basic geochemical understanding of the IRW environment. Many of these 100-plus so-called different constituents were actually just different names for the same method for the same constituent. Therefore, these data needed to be combined in the PCA. Also, many other constituents were analyzed for but were never detected above laboratory detection limits. These are certainly not important constituents in the IRW environment and they therefore provide no useful information to the PCA. Defendants' experts Drs. Cowan, Johnson, and Murphy all lack the fundamental geochemical knowledge necessary for proper consideration of these issues. I will provide further discussion of these and related points later in this declaration.

13. After Dr. Olsen extracted pertinent data from the project database, the next step in his PCA methodology was reduction of the extracted data into PCA datasets. In environmental science, "reduction" refers to the evaluating, grouping, retaining, eliminating, combining, and formatting of data in order to obtain a dataset usable for

subsequent analysis (also commonly referred to as "data treatment"). Data reduction (treatment) is a part of the DQO process. It is a necessary step, followed in all environmental studies. Dr. Olsen established a set of standard protocols to follow that are detailed in his expert report in Section 6.11.2 ("Steps of PCA"), and in particular under Step 6 ("Preparing Data for PCA") and Step 8 ("Identifying Parameters that Meet PCA Criteria"). To document the standard protocols he followed, Dr. Olsen placed the data extracted from the project database into two Microsoft EXCEL workbook files named "PCA_Main_Database_Water.xls" and "PCA_Main_Database_Solids.xls". He then carried out all of the data treatment steps within these two files. The two files (containing the data treatment documentation) were provided in his considered materials. This is important to note because Defendant's expert Dr. Cowan argues that Dr. Olsen's data treatment process is not reproducible. I have thoroughly examined the steps that Dr. Cowan took in attempting to reproduce the data treatment process for surface water samples. My examination confirmed that Dr. Cowan failed to follow the correct protocols as detailed and documented by Dr. Olsen. Instead, Dr. Cowan ignored these protocols and as a result he made many mistakes. I have documented the mistakes Dr. Cowan made in my previous declaration (Attachment A). Since Dr. Cowan did not follow Dr. Olsen's data treatment procedure, his claim that it is not reproducible is baseless. Since Dr. Cowan was unable to reproduce the data treatment procedure (because he did not follow the correct protocols), his conclusion that Dr. Olsen must have somehow "hand-selected" the data is equally baseless. It would be impossible for Dr. Olsen to follow a set of standard protocols (which he did) and at the same time hand-select the data.

14. As discussed above, Dr. Olsen followed the DQO process to ensure that the appropriate type, quantity, and quality of data were collected for PCA. Dr. Olsen included as many samples and variables (constituents) as possible in order to provide the most comprehensive PCA, one that best represented environmental conditions in the IRW. I have compared the numbers of sampling locations included by Dr. Olsen in the IRW PCA with the numbers of sampling locations included in comparable watershed studies where PCA was also employed. In all cases, the number of sampling locations included in the IRW PCA far exceeded those in the other watershed studies. The larger number of sampling locations means that the intensity of spatial coverage (the sampling density) was enhanced considerably compared to a typical watershed study. I have also compared the types and numbers of variables (chemical/bacterial constituents) used in the IRW PCA with those used in other watershed studies. I find the types and numbers of variables to be at least comparable. In many cases, the numbers of variables included in the IRW PCA exceed those in a typical watershed PCA study. Furthermore, as an environmental scientist and geochemist, I find that Dr. Olsen has included all variables present naturally in a watershed environment along with all variables potentially introduced to the IRW by sources of contamination (many of these are the same variable). This is important because Defendants' expert Dr. Cowan (and to a lesser degree Dr. Johnson) argue that the numbers of samples and variables are inadequate. On the contrary, the numbers of samples included in the IRW PCA are more than enough. The sampling density is more than adequate to represent the IRW environment. And no important variables have been excluded. Again, all chemical and bacterial constituents of importance to characterizing the IRW environment (both those present naturally and

those resulting from contamination) were included in the IRW PCA. As I discussed previously, Defendants' experts all lack the fundamental geochemical background necessary to properly evaluate these issues. Dr. Cowan doesn't know that TKN (total Kjeldahl nitrogen) is a measure of organic nitrogen plus ammonia. Dr. Johnson doesn't know the formula of the orthophosphate anion (PO_4^{3-}), one of the most important forms of phosphorus in the IRW environment. Knowledge of the chemical composition of a watershed environment and the geochemical behavior of the chemical and bacterial constituents is very important. Because the Defendants' experts lack this knowledge, they cannot properly evaluate which constituents should be included in the PCA.

15. Dr. Olsen sought to use the greatest number of samples possible in the PCA in order to best represent environmental conditions in the IRW. He accomplished this objective by allowing the inclusion of samples with a small number of missing or un-analyzed constituents. For example, in one of his surface water PCA datasets called "SW3", there were 573 samples and 26 variables (chemical/bacterial constituents) for a total of 14,898 possible data values. Of these, 915 data values (about 6% of the total) were missing or blank (primarily due to bacterial results that were rejected for quality control reasons or samples that were not analyzed for bacteria). The protocol that Dr. Olsen followed was that a sample had to be complete for at least 20 of the 26 variables in order to include it in this particular dataset (SW3). This is important because Defendants' experts Cowan and Johnson have criticized the inclusion of samples with missing values, claiming that this leads to bias or distortion. There are a number of related points that are important to note about this missing value issue and I will address all of them in turn in this and following paragraphs. First, Dr. Olsen did not have to

include samples with a small percentage of missing values in order to conduct a PCA and obtain representative results for the IRW. In fact, he conducted a surface water PCA identical to SW3 in every respect except that it did not include samples with missing values. This PCA run was called "SW15". It contained 267 samples and 26 variables for a total of 6,942 data values, none of which were missing. This SW15 PCA was certainly large enough for Dr. Olsen to characterize contaminant source patterns in the IRW environment. However, there is considerable usable information contained in the remaining 306 samples with between 20 and 25 variables, and therefore Dr. Olsen sought to use this information to the maximum extent possible. So what Dr. Olsen did was he ran the PCA both ways (with and without the missing values). He then compared the PCA results between the two runs (SW3 and SW15) to determine whether the results were comparable, i.e., whether or not the PCA was sensitive to the inclusion of samples with a small percentage of missing values. Such an analysis is termed a "sensitivity analysis" and is a standard practice in environmental science. A sensitivity analysis is essentially a type of test to evaluate the degree of error in an analysis. Dr. Olsen conducted many different sensitivity analyses (22 different tests for the surface water PCA alone). He provided the results of all sensitivity analyses in his considered materials. In the SW3/SW15 case, he compared the two results and determined that the patterns they exhibited were similar (in fact, nearly identical). Hence he was able to conclude that the PCA was insensitive to the inclusion of samples with a small percentage of missing values. It is important to point out that all three Defendants' experts (Drs. Cowan, Johnson, and Murphy) ignored this and other sensitivity analyses conducted by Dr. Olsen. In other words, they make their claim of "bias" without even considering or evaluating the results of the sensitivity analysis. Furthermore, they make this claim without conducting any other valid tests to support it (as discussed below, Dr. Cowan conducted a different type of test, but it was flawed).

16. Dr. Olsen applied a recognized method in his PCA designed to accommodate the inclusion of samples with a small number of missing values. This method is called pair-wise deletion. In pair-wise deletion, the correlation coefficient between two variables is calculated after excluding samples with one or both values missing. Pair-wise deletion is a standard method available in many computer packages that conduct PCA, including the program called SYSTAT used by Dr. Olsen (see Attachment B for a description of this method as provided in the SYSTAT manual). Defendant's expert Dr. Cowan apparently did not understand that pair-wise deletion was appropriately used by Dr. Olsen. Indeed, he confused it with another method known as mean substitution. In mean substitution, the correlation coefficients between two variables are calculated after substituting the means for all samples with no missing values for the missing values. Dr. Cowan claims that these two methods (pair-wise deletion and mean substitution) are equivalent, but in fact they are not equivalent. They do not result in the same PCA output. This explains why the two methods are offered as alternative methods in many statistical programs that conduct PCA (for example, see Attachment C). Dr. Olsen appropriately used pair-wise deletion for the IRW PCA. Furthermore, he tested the pair-wise deletion method of handling the small percentage of missing values via sensitivity analysis. As discussed earlier, Dr. Olsen conducted a PCA on surface water dataset SW3, which contained missing values and was processed using pair-wise deletion. Concurrently, he conducted a PCA on surface water dataset SW15,

which was the same as SW3 but with samples containing missing values removed (this is the same as conducting PCA on SW3 but without employing the pair-wise deletion method). He then compared the PCA results produced by the two datasets. So his comparison was in effect a test of the pair-wise deletion method. Dr. Olsen's examination of the two PCA results revealed no substantial difference between them, thus confirming that the pair-wise deletion method was valid. Again, this is important because Defendant's expert Dr. Cowan (ignoring the sensitivity analysis) claimed a "distorting effect". Dr. Cowan tried to demonstrate his claimed distorting effect by conducting a type of missing data analysis. However, his method was flawed because he failed to log-transform the data, he improperly grouped the data, and he used inappropriate statistics. What I mean by inappropriate statistics is that he used the minimum, maximum, and mean values in his analysis, whereas he should have used the correlation coefficient. I have redone his analysis using the correct log-transformation, the proper data grouping, and the correlation coefficient. I found that the average difference between the correlation coefficient resulting from use/non-use of the pair-wise deletion method was 0.023. This value is very small and indicates no perceptible bias. This is in agreement with the results of the sensitivity analysis.

17. The IRW investigation, like any environmental investigation, included the collection of field duplicate samples (the regular sample is homogenized and split into two equivalent portions). These duplicate samples (also called splits or replicates) were submitted blind to the analytical laboratory along with the regular samples (i.e., multiple analysis of the same sample for the same set of constituents). Only a small percentage of the regular samples have associated blind duplicates. These are randomly selected over

the course of the investigation. The purpose of the blind duplicates is solely to assess the precision of the laboratory for the analytical constituents. This is a component of the DQO process discussed earlier. The concern is to evaluate whether the laboratory is producing analytical data of sufficient quality. This is important because Defendant's expert Dr. Cowan has completely misinterpreted what these duplicate data represent and how they should be used in the PCA. Dr. Cowan argues that the duplicates should be treated as separate samples in the PCA. But they are not separate samples. They are the same sample, just analyzed twice. To include both the duplicate and the regular sample in the PCA would therefore have been inappropriate. Dr. Olsen did not and should not have included these duplicates in the PCA as if they were separate samples. Instead, he employs the correct approach and averages (takes the mean value of) the duplicates with the regular samples prior to the PCA. Defendant's expert Dr. Murphy also averages the duplicates in his PCA (see page 30 of Murphy's rebuttal report). Renowned statistician John C. Davis states the following in his textbook: "In geochemical analyses, it is common practice to make multiple determinations, or replicates, of a single sample. The most nearly correct analytical value is taken to be the mean of the determinations." (Page 35 in "Statistics and Data Analysis in Geology, Third Edition").

18. In a watershed investigation such as the IRW, the data collected will inherently contain results that are reported by the laboratory as below their analytical detection limits. Such data are referred to as nondetect or left-censored (meaning that the true value is between zero and the detection limit). This occurs when the analytical instrument employed is not sensitive enough to measure the concentration with certainty. However, it does not mean that the true concentration is zero. In environmental science,

a common approach for handling such left-censored data is to substitute the values reported as below the detection limit with one-half of the detection limit (the mid-point between zero and the detection limit). Dr. Olsen followed this approach with left-censored data for purposes of his PCA. Defendant's expert Dr. Murphy also followed this approach (see page 30 of Dr. Murphy's rebuttal report). Defendant's expert Dr. Johnson also follows this approach, e.g., he states on page 1172 of a scientific paper: "Data reported as non-detect were represented in the dataset by one-half the value of the sample detection limit." ("Identification of historical PCDD/F sources in Newark Bay Estuary subsurface sediments using polytopic vector analysis and radioisotope dating techniques, Chemosphere, Vol. 36, No. 6, pp. 1167-1185, 1998"). Defendant's expert Dr. Cowan surprisingly criticizes this approach and argues that zeros should be used instead. In fact, only in cases where a constituent is a synthetic contaminant or not naturally-occurring (such as a PCB or chlorinated solvent) would it be appropriate to consider substituting a nondetect with a zero. In cases where the constituent is naturally-occurring the value cannot be zero, and therefore the zero substitution should not be used.

19. In the IRW investigation, Dr. Olsen sought to make use of surface water samples routinely collected in the IRW by the United States Geological Survey (USGS). Dr. Olsen incorporated the USGS samples into his overall sampling design. He collaborated with the USGS to have their samples analyzed for the same suite of constituents required for the PCA and according to the same DQO process. The USGS samples were collected at monitoring stations located along major stream segments in the IRW, whereas the other samples collected during the IRW investigation were located primarily along smaller stream segments. Hence the USGS samples and the other

samples represent two different portions of the overall IRW surface water environment. Defendant's expert Dr. Cowan conducted an evaluation comparing the USGS data groups and the other (non-USGS) data groups as if they both represented the same environment. In fact, they do not. Dr. Cowan conducted a PCA separately on the USGS and the non-USGS portions of the SW3 dataset, showing that the results (as expected) were different. He argues that this difference indicates an incompatibility between the USGS and non-USGS datasets. In fact, his argument is totally illogical because these two datasets would not be expected to be comparable in the first place.

20. In extracting data from the master database and reducing the extracted data to subsets for PCA, Dr. Olsen followed a set of documented protocols designed to obtain comprehensive and scientifically valid and reliable data. These protocols are described in detail in Section 6.11 of his expert report. They include the handling of duplicate sample data (also discussed above), the handling of constituents analyzed by various different analytical methods (another type of quality control assessment), and the handling of data qualified as rejected based on quality control evaluations. Dr. Olsen produced files in his considered material documenting his data handling practices at every step along the way. Defendants' expert Dr. Cowan attempted to reproduce Dr. Olsen's SW3 PCA dataset, starting with extraction of the data from the project database. However, Dr. Cowan was unable to reproduce Dr. Olsen's surface water (SW3) PCA dataset because he failed to extract all relevant data from the project database, and because he apparently lacked the understanding necessary to follow Dr. Olsen's protocols. I have documented all of the errors Dr. Cowan made in attempting to reproduce Dr. Olsen's SW3 PCA dataset in my previous declaration (see Attachment A).

concern and proceeded to conduct a PCA on a surface water dataset that contained 60 variables, many of which had very high percentages of left-censored data (some variables consisted entirely of nondetects). Most of these highly left-censored variables were also multiply left-censored. So in cases where Dr. Cowan observed that such a highly and multiply left-censored variable was important in the PCA, it in fact was not really important. In fact, such a variable's observed importance (by Dr. Cowan) was due to an artificial variance resulting solely from the presence of multiple detection limits. Dr. Olsen eliminated this artificial variance by following his documented protocol. Dr. Cowan did not. So when Dr. Cowan claims that inclusion of 60 variables in the surface water PCA (as opposed to Dr. Olsen's 26 variables in SW3) causes Dr. Olsen's poultry signature to "disappear", what he really should be stating is that he can artificially obscure the poultry signature by including highly and multiply left-censored (i.e., very inappropriate) data. It is important to also note that even given the obvious attempt by Dr. Cowan to artificially obscure the poultry signature, he was unable to completely do so. I have carefully examined Dr. Cowan's 60 variable PCA results. For the most part, the pattern shaped by the impact of poultry waste in the IRW remains even in the face of Dr. Cowan's attempt to artificially mask it.

If Dr. Cowan would have followed the correct and documented protocols, he would have exactly reproduced Dr. Olsen's SW3 PCA dataset.

21. As discussed previously, analytical results for a constituent may be reported by the laboratory as below their analytical detection limits (i.e., left-censored or nondetect). In many environmental investigations, especially when the investigation is a large watershed study conducted over multiple years (like the IRW), it is common to have multiple detection limits for a particular constituent. This happens for a variety of reasons. For example, the laboratory may change their instrumentation and the detection limit may change accordingly. Or differences in the overall chemical compositions of various samples may cause the detection limits for a particular analyte to fluctuate between samples. In environmental science, datasets that contain left-censored (below detection limit or nondetect) values for a particular constituent but with more that one detection limit are referred to as "multiply left-censored" datasets. Care must be exercised when applying statistical analyses to such multiply left-censored datasets. This is particularly important for statistical analyses that involve calculation of variances (such as PCA). The concern is that if a high percentage of the variable (constituent) is nondetect and also contains many different detection limits, the variance may be unduly influenced by the detection limits and not representative of the true (though unknown, because they are below detection limits) values. To eliminate this possible undue influence, Dr. Olsen followed a documented protocol whereby variables with high percentages of nondetects (especially if they were multiply left-censored) were excluded from his PCA datasets. This is important to note because Defendant's expert Dr. Cowan expressed concern over this issue. Surprisingly, however, Dr. Cowan ignored his own

I declare under penalty of perjury, under the laws of the United States of America, that the foregoing is true and correct.

Executed on the 5rd day of June, 2009.

Richard W. Chappell

Attachment A

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF OKLAHOMA

STATE OF OKLAHOMA,	)	
	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 05-cv-329-GKF(PJC)**
	)	
TYSON FOODS, INC., et al.,	)	
	)	
Defendants.	)	

DECLARATION OF RICK W. CHAPPELL, Ph.D.

I, Rick W. Chappell, Ph.D., hereby declare as follows:

A. BACKGROUND

1. Since June 2005, I have been an independent contractor and owner of Environmental Science Solutions LLC, an environmental consulting company. My educational background includes a Bachelor of Science degree with special distinction in Geology from the University of Southern Colorado in Pueblo, Colorado, in 1980 and a Doctor of Philosophy degree in Geochemistry from the Colorado School of Mines in Golden, Colorado, in 1989.

2. Since 1997, I have been an adjunct professor and instructor in the Civil Engineering Department at the University of Colorado at Denver and Health Sciences Center where I have taught courses in water quality laboratory and environmental engineering. From 1985 to 2005, I was an environmental scientist at Camp Dresser & McKee Inc. ("CDM") in Denver, Colorado, where I was a senior geochemist and environmental statistics discipline leader. I have extensive experience in performing environmental investigations and studies, evaluating the environmental fate and transport

Attachment A

of chemicals in the environment, determining the cause or source of contamination in the environment, and conducting statistical analyses related to environmental contamination.

3. In November 2004, CDM was retained by the Oklahoma Attorney General to perform an investigation concerning environmental contamination found in the Illinois River Watershed ("IRW"). I have been involved in this study, first with CDM and then as a subcontractor to CDM, since 2005. My primary role on the IRW project has been to support CDM staff and to conduct statistical analyses. In this capacity, I have helped plan and direct various aspects of the investigation of the environmental contamination found in the IRW, including development of sampling designs, chemical analyses, and various statistical analyses such as the principal components analysis (PCA).

B. Opinions of Charles D. Cowan, Ph.D.

4. I have reviewed the opinions of Charles D. Cowan contained in his expert rebuttal report and deposition testimony. In my opinion, Dr. Cowan lacks relevant experience and specialized knowledge in environmental science and agricultural non-point source pollution necessary to offer opinions regarding contamination in the IRW. This is evident in his November 2008 Rebuttal Report where he summarizes his education and experience (Paragraphs 1 and 2, Page 2; Appendix 1, Pages 61-62), presents his resume (Appendix 2, Pages 63-68), including publications, and provides a listing of his past litigation experience (Appendix 3, Pages 69-72), which confirms that his analytical and statistical experience is limited to economic analysis, demographics, and census surveys, i.e., social science.

5. In his Deposition Testimony, Dr. Cowan did identify a few cases (not listed in his Rebuttal Report) where he was involved with environmental studies.

However, his involvement in these studies was again limited to economic or demographic analysis and he himself had no direct involvement with the core environmental analysis of contaminant sources or the nature and extent of contamination, which were conducted by other experts. It is clear that Dr. Cowan has never actually conducted an environmental analysis or critically reviewed the results of an environmental analysis conducted by others. His lack of experience in the area of environmental science should disqualify him from making any considered or relevant opinions in this (the IRW) case. His lack of experience in environmental science is perhaps best summed up by his own deposition testimony:

Deposition Transcript, Vol. I, Page 7:

> **Q Have you had any cases where you've actually done an environmental analysis as an expert?** **09:13AM**
> A No.

6. Dr. Cowan attempts to minimize his lack of experience by stating in his Rebuttal Report and Deposition Testimony that he was not tasked to render an opinion specifically regarding contamination in the IRW, but rather that he was only tasked to render an opinion regarding the ability of Dr. Olsen to draw conclusions from his principal components analysis (PCA), one of several scientific approaches used by Dr. Olsen and other experts in evaluating contaminant impacts in the IRW. Dr. Cowan would like the court to believe that knowledge or experience in environmental science (or any field within the physical sciences) is not required, i.e., that "analysis is analysis" and "statistics is statistics" essentially regardless of the scientific context to which it is applied. While there is some strict mathematical truth to this, any good scientist will acknowledge that one needs to have a solid connection between the analysis/statistics

applied and the science it is applied to; that one should not conduct such an analysis in a vacuum; and, in particular, that a fundamental understanding of the variables (analytes or contaminants) and their behavior in the environment is requisite to drawing conclusions. In addition, one cannot evaluate the relevance and reliability of a statistical analysis without considering other scientific evidence, such as the evidence contained in the other IRW studies, which of course Dr. Cowan did not review or even consider (ostensibly because he was not tasked to do so, though in reality he is unqualified to do so). Environmental variables like phosphorus and bacteria are totally different than economic or demographic variables, and it is naïve and dangerous to assume otherwise. Dr. Cowan's qualifications clearly show that he lacks the necessary connection and fundamental understanding necessary to adequately review the work of an experienced environmental scientist, and therefore his opinions regarding Dr. Olsen's PCA are misleading, irrelevant, and unreliable. This lack of experience, specifically with regard to PCA, is perhaps best summarized by his own deposition testimony:

Deposition Transcript, Vol. I, Page 69:

Q So has your work in the -- with PCA been primarily involving studies within the social sciences? 10:41AM
A Yes.

Deposition Transcript, Vol. I, Page 70:

Q Have you done any kind of PC analysis with a dataset similar to Dr. Olsen's? 10:42AM
A No.

Further, from his Deposition Testimony, it is clear that Dr. Cowan has no specific knowledge regarding the IRW. He has no concept even of the nature of the study area:

Deposition Transcript, Vol. I, Page 48:

Q Have you been to any location within the
Illinois River watershed?
A Not that I can think of or name.
Q Have you taken any car trips, for example,
that would show you the Illinois River or Lake **10:05AM**
Tenkiller?
A No.
Q Have you looked at any streams that might be
within the Illinois River watershed?
A No. 10:05AM
Q Have you looked at any areas where there are
chicken houses in the Illinois River watershed?
A No, sir.

And he has no concept or fundamental understanding of the nature of contaminant transport, either potentially within the IRW or in general:

Deposition Transcript, Vol. I, Page 51:

Q Thank you. What do you know about the **10:16AM**
hydrology of the IRW?
A I can't claim to have any specific knowledge
of the hydrology.
Q Do you have any understanding of how
contaminants move in the environment of the IRW? **10:17AM**
A I haven't studied that.

7. Dr. Cowan's lack of experience with environmental science and environmental data has led him to conduct an evaluation of Dr. Olsen's PCA in mathematical isolation, i.e., completely insulated from the core scientific foundations it has been applied to. This has not only lead him to express irrelevant opinions but also to express false opinions based on so-called "key problems" that he identifies in his Rebuttal Report, but which in fact are not really present. There are approximately eight key problems that Dr. Cowan identifies, but they all essentially fall within two general categories:

(a) Extraction of data from the main project database and PCA dataset reduction.

(b) Data processing and the mechanics of the PCA method.

All of the key problems identified by Dr. Cowan result, directly or indirectly, from his lack of experience with environmental data and his lack of a fundamental understanding of basic environmental phenomena. Dr. Cowan criticizes the sampling, retaining, or eliminating of certain environmental variables without having any understanding or appreciation of what these variables are, how they behave in a complex environmental system like the IRW, or even the fundamental principals of geochemistry, hydrology, and contaminant transport mechanisms involved. In short, Dr. Cowan lacks the fundamental and specific knowledge necessary to make any opinions regarding the data collected or the variables retained in the PCA dataset for purposes of environmental analysis. Therefore, his opinions and testimony in this regard are irrelevant and unreliable.

8. One of Dr. Cowan's major criticisms or "key problems" is that, starting with the main IRW project database, he was unable to reproduce precisely the reduced datasets used by Dr. Olsen in his PCA. Even though the procedures followed by Dr.

Olsen are provided in his expert report, and documented in detail in intermediary files contained in his considered materials, Dr. Cowan did not examine these procedures properly, chose to not follow them, or he made gross mistakes in doing so. One of the fundamental mistakes Dr. Cowan made was that he did not extract all of the appropriate data due to his use of an incorrect database query. In particular, Dr. Cowan makes the following statement in his Rebuttal Report:

> Page 35, Paragraph 80:
> "We attempted to reproduce the values in the SW3 Excel sheet and the PCA values in Appendix F of the CDM report. All of the records from the master database with "SW:S" in the sample groups were downloaded into an Excel file. This download produced an Excel sheet with all of the surface water data."

This statement does not fit the facts, in that the "download" conducted by Dr. Cowan would NOT have extracted all of the surface water data contained in the master database. This illustrates one of several fundamental mistakes make by Dr. Cowan. His use of the "SW:S" criterion caused him to miss hundreds of data records corresponding to sample splits or replicate analyses that were used as averages by Dr. Olsen in his PCA. A competent environmental scientist would have understood and realized that such splits/replicates are common, would have recognized them in the database, and would have ensured that they were also extracted. Furthermore, an environmental scientist in the process of reproducing a dataset would have immediately recognized that the different values obtained for a percentage of the samples could be the result of mishandling of such split/replicate data. Dr. Cowan did not recognize this because he has no experience with environmental data or environmental databases; it seems evident that he was unaware even that such splits/replicates were present in the IRW database (even though this is documented in detail) and that they would be expected to be present in any large environmental study of this type. As a result, Dr. Cowan observes that a certain percentage of the data have different values between his and Dr. Olsen's PCA dataset, but he has no fundamental experience or basis to question whether these "discrepancies" could be due to the averaging of splits/replicates, rather he makes the ridiculous assumption or false insinuation that Dr. Olsen must somehow be changing the data. This demonstrates gross lack of relevant experience.

9. Dr. Cowan failed to recognize, examine, or use a specific protocol followed by Dr. Olsen, and documented in detail, to obtain the most reliable phosphorus data available. To illustrate this mistake, Attachment A is a screen-shot of a portion of one of Dr. Cowan's extracted PCA datasets (obtained from his considered materials). The important point to note here is that the highlighted portion (intersection of row 88 and column U) represents a particular sample for which Dr. Cowan's dataset does not have a corresponding value for Total Phosphorus (indicated as P_T), i.e., the cell at the intersection is blank. The number that appears above "P_T" is "149", which is the parameter key code for one of several analytical methods used in the IRW for this particular variable. In using only this parameter key code for Total Phosphorus, Dr. Cowan failed to extract critical data from the project database (due to his failure to follow Dr. Olsen's protocol). Hence Dr. Olsen's PCA dataset did have a corresponding value for Total Phosphorus for this sample (the value is actually 0.17 mg/L which corresponds to parameter key code 147). This example illustrates just one of hundreds of cases where Dr. Cowan's PCA dataset contained a missing value when in fact there was available data. Dr. Cowan makes the following statement in his Rebuttal Report:

Page 38, Paragraph 85:

> "In addition, there are 499 additional values that were missing data in the Access database, but which suddenly have data in Dr. Olsen's analysis file."

As discussed above, this statement certainly does not fit the facts. Data does in fact exist for these 499 "missing" values, it's just that Dr. Cowan failed to extract it because he did not follow or did not understand Dr. Olsen's protocol. Any competent environmental scientist would have readily recognized this "key problem" and quickly uncovered its cause; furthermore, any competent environmental scientist would have been sensitive to the fact that phosphorus is a very important contaminant variable in the IRW and that data for it cannot just "suddenly" appear. Dr. Cowan missed an important step in the data reduction process due to his lack of experience and/or carelessness.

10. Dr. Cowan failed to recognize that certain data were rejected due to quality control issues, even though this was documented in detail, and even though such rejected data were clearly indicated with an "R" qualifier in the "ValidFlag" field and a "TRUE" qualifier in the "Rejected" field in the master database. As a result, Dr. Cowan ended up using these rejected data in his PCA dataset. This primarily affected the bacteria data. With regard to this issue, Dr. Cowan makes the following statement in his Rebuttal Report:

Paragraph 85, Sentence 1:

> "To summarize…the remaining 66 were decreed by Dr. Olsen to be missing when they in fact had data."

Again, this statement does not fit the facts. These data were "missing" in Dr. Olsen's PCA dataset because they were rejected due to quality control problems. Attachment B is a screen-shot of one of Dr. Cowan's extracted PCA datasets. Note that the highlighted portion (intersection of row 450 and column M) shows a value of 0.5 (organisms/100 ml) for Fecal Coliform bacteria for a particular sample in Dr. Cowan's dataset. In Dr. Olsen's corresponding dataset, as noted by Dr. Cowan, this value was missing, i.e., blank. Attachment C is a screen-shot of the main PCA database file, which contains data extracted directly from the main project database, and which clearly shows that this datum was rejected and provides the explanation as to why it was not used in Dr. Olsen's PCA. Note that the reason the value appears as 1 org/100 ml instead of 0.5 org/100 ml (intersection of row 17733 and column M) is because 1 org/100 ml was the laboratory detection limit and the result was below the detection limit, and therefore Dr. Cowan used one-half of this value (0.5) in his PCA dataset (following the protocol used by Dr. Olsen). This information concerning why the datum was not used by Dr. Olsen was made available to Dr. Cowan but he chose to ignore it. Any competent environmental scientist would have easily tracked it down and understood the root issue.

11. As demonstrated above, Dr. Cowan's lack of experience with environmental data has led him to use an incorrect dataset. Therefore, all of his "key problem" assertions with regard to extraction of data from the main project database and PCA dataset reduction are in fact false. Furthermore, since he has used an incorrect dataset, all of his independent analyses using that dataset, as provided in his Rebuttal Report, are irrelevant and unreliable.

Dr. Cowan's lack of experience in environmental science also translates into his irrelevant and unreliable evaluation of Dr. Olsen's application of PCA with respect to the IRW data. In addition, Dr. Cowan made mistakes even in interpreting or understanding the mechanics of the PCA methodology used by Dr. Olsen. For examples, though this was stated precisely in Dr. Olsen's expert report, Dr. Cowan did not understand Dr.

Olsen's handling of missing data values, as indicated by the following statement in his Rebuttal Report:

> Paragraph 44, Page 19:
>
> "As noted above, not all samples have measurements on all observations. In fact, there is a very significant amount of missing data. Dr. Olsen disguises this by substituting for the missing data. He plugs in the mean of a variable for the actual (though missing) value. Only 267 of the 573 samples used by Dr. Olsen have complete data. This means only 47% – less than half – of the observations have real data actually observed in the field. This means that more than half of Dr. Olsen's observations have data that Dr. Olsen substituted rather than real data."

This statement does not fit the facts. First, mean values were NOT substituted for missing data. Dr. Olsen used a method called pair-wise deletion to handle missing data in his PCA, followed by substitution of mean z-scores (which have values of zero) to calculate PC scores. The fact that Dr. Cowan was unfamiliar with this method, or that he was unable to distinguish it from "mean substitution", brings into question his actual experience with PCA, both in general and with regard to its application in environmental science. Second, Dr. Cowan is intentionally misleading the court to believe that half of the data used in Dr. Olsen's PCA was missing, whereas in fact only a small percentage (about 6%) of the values were missing in the dataset he is referring to in his statement. Finally, Dr. Cowan ignored or did not understand the results of the extensive sensitivity analyses conducted by Dr. Olsen, which included several PCA runs where missing values were not allowed in various datasets.

12. Another area where Dr. Cowan demonstrates his lack of experience with environmental data concerns the handling of values that were reported as below a laboratory detection limit. In his Rebuttal Report, he criticizes several aspects of this issue. First, he criticizes Dr. Olsen's substitution of one-half the detection limit for these nondetect values (the midpoint between zero and the detection limit):

Paragraph 57, Page 26:

> "In the data analyzed by Dr. Olsen, he also has a number of values that are non-detects, meaning the measurement method used by the researchers cannot measure any trace measure of a chemical or organic value. Rather than treat this as a zero (not detected), Dr. Olsen substitutes the midpoint between zero and the detect limit for a chemical."

Here, Dr. Cowan is naïve in thinking that a value reported as below laboratory detection limits means that the value is zero, i.e., that the variable is not present. An environmental scientist (even a first-year graduate student) would immediately recognize that all of the variables included in Dr. Olsen's PCA represent naturally occurring chemicals or biological organisms present in the environment, and that they are present in all water and soil media in the IRW, albeit at concentrations below laboratory detection limits in certain samples. Such environmental variables are totally different than the sociological variables Dr. Cowan is used to dealing with where one might expect certain variables to actually have zero values. Dr. Cowan's statement also demonstrates his lack of experience in applying PCA to environmental data, where it is a very common practice to substitute one-half the detection limit for nondetects. In his Deposition Testimony, Dr. Cowan changed his mind about the appropriateness of this practice (after acknowledging that in the interim between his Report and Testimony he actually reviewed similar published studies):

Deposition Transcript, Vol. I, Page 132:

Q Okay. Now, you testified before lunch I believe that you're not criticizing Dr. Olsen by using the midpoint between zero and the detection limit when he ran his PCA, correct, for non-detects? **01:34PM**

A I agree, I am not criticizing him for not using zero. Using the midpoint between zero and the lower limit of the detection level is an acceptable procedure. 01:35PM

This represents another contradiction underscoring his ignorance with the application of PCA to environmental data. Second, he suggests in his Rebuttal Report that using one-half the detection limit for nondetects versus some other protocol, such as closer to zero or closer to the detection limit, followed by log-transformation (the approach Dr. Olsen used and Dr. Cowan seems to agree is applicable), would "have huge effects on the outcome" (Paragraph 74, Page 32) due primarily to variable detection limits. Dr. Olsen sought to minimize this possible effect by limiting his PCA to only those variables with very low percentages of nondetects, an approach that is documented in Dr. Olsen's expert report but never fully understood, recognized, or evaluated by Dr. Cowan. In fact, Dr. Cowan has performed no testing of any kind to support his contention, and therefore his contention is unreliable. Finally, in a totally absurd fashion, Dr. Cowan ignores his own concern and performs a PCA on a dataset that contains variables with very high percentages of nondetects, i.e., an extended set of 60 variables where for many of the variables the data were all or nearly all nondetects. The statements summarizing the results of his analysis appear at the end of the section of his Rebuttal Report concerning Dr. Olsen's PCA work:

Paragraph 114, Page 50-51:

> "These PCA results indicate that other variables are important in addition to the 26 variables discussed in the CDM report, and that the "signature" discovered by Dr. Olsen disappears when he brings in the full set of data available."

Any experienced environmental scientist would immediately recognize Dr. Cowan's contradiction here and dismiss his results as irrelevant and unreliable. All of Dr. Cowan's opinions in this case are irrelevant and unreliable. He is not qualified.

I declare under penalty of perjury, under the laws of the United States of America, that the foregoing is true and correct.

Executed on the 18th day of May, 2009.

Richard W. Chappell

Richard W. Chappell

Attachment A

Cowan016488_usgscompare_rwc.xls

	A	S	T	U	V	W	X	Y
1	sample group	42	144	149	143	85	71	
2	EDA_Sample	NO2_NO3	P_SOL_REAC	P_T	P_TD	SO4	TDS	TKN
77	EOF-SPREAD068:6/18/2006:SW:S:-:-	0.05	0.039	0.772	0.081	6.31	166	
78	EOF-SPREAD071:5/10/2006:SW:S:-:-	0.05	0.991	1.06	1.02	2.12	71	
79	EOF-SPREAD073E:6/22/2006:SW:S:-:-	2.36	0.139	0.274	0.165	5.9	170	
80	EOF-ZPEOF001:4/25/2006:SW:S:-:-	0.541	0.526	1.12	0.582	6.29	154	
81	EOF-ZPEOF030:4/25/2006:SW:S:-:-	0.05	0.282	0.453	0.343	21	140	
82	HFS-02:4/29/2006:SW:S:-:INITIAL	1.93	0.056	0.078	0.07	3.98	288	
83	HFS-02:5/1/2006:SW:S:-:-	2.4	0.189	0.248	0.224	5.86	202	
84	HFS-02:5/10/2006:SW:S:-:-	3.09	0.141	0.179	0.162	7.35	137	
85	HFS-02:5/11/2006:SW:S:-:PEAK		0.179	0.229	0.182			
86	HFS-02:5/4/2006:SW:S:-:PEAK	2.34	0.183	0.193	0.19	5.25	210	
87	HFS-02:5/6/2006:SW:S:-:-	3.46	0.067	0.099	0.075	7.01	168	
88	HFS-02:6/15/2005:SW:S:-:-	2.297				3.66	142	
89	HFS-02:6/15/2006:SW:S:-:BF1	2.2	0.01	0.034	0.015	5.27	156	
90	HFS-02:6/27/2005:SW:S:-:-	2.615				4.45	159	
91	HFS-02:6/8/2006:SW:S:-:-	2.36	0.039	0.055	0.042	5.23	162	
92	HFS-02:7/11/2005:SW:S:-:-	2.421	0.020937964	0.023516357	0.021903833	4.2	147	
93	HFS-02:7/11/2005:SW:S:-:BF1	2.699	0.032	0.032644026	0.030643253	3.08	213	
94	HFS-02:8/1/2006:SW:S:-:BF2	2.29	0.046206326	0.050911447	0.047289504	4.24	199	
95	HFS-02:8/27/2005:SW:S:-:BF2	2.698	0.018306764	0.023337581	0.021171053	4.16	152	
96	HFS-02:9/16/2005:SW:S:-:-	2.398				3.74	135	
97	HFS-04:3/10/2006:SW:S:-:-	9.57	1.682244844	1.951863692	1.895390079	41.52	316	
98	HFS-04:3/10/2006:SW:S:-:PEAK	8.53	1.499471179	1.831536566	1.695894861	40.35	309	

data1 / data2 / dataones / datazeroes / pcaruns / Sheet1

Attachment B

Cowan016488_usgscompare_rwc.xls

	A	I	J	K	L	M	N	O	P	
1	sample group	69	100	103	72	39	81	75	76	
2	EDA_Sample	CU_T	ECOLI	ENTERO	FE_T	FECAL	K_T	MG_T	MN_T	NA_T
437	USGS-07195500:10/4/2007:SW:S:-:-	0.0021	1100	7600	0.912	1800	4.82	1.79	0.0917	
438	USGS-07195500:12/12/2007:SW:S:-:-	0.0012	1100	3300	0.294	1100	4.5	1.88	0.0404	
439	USGS-07195500:12/6/2007:SW:S:-:-	0.0012	0	1	0.102	1	5.2	2.27	0.0234	
440	USGS-07195500:12/7/2006:SW:S:-:-	0.001	140	460	0.168	140	3.76	2.39	0.0241	
441	USGS-07195500:2/7/2007:SW:S:-:-	0.0011	2	5	0.126	2	3.82	2.27	0.0184	
442	USGS-07195500:3/10/2006:SW:S:-:-	0.0012	140	180	0.208	140	6.6	2.1	0.0384	
443	USGS-07195500:3/9/2006:SW:S:-:-	0.0011			0.187		6	2.09	0.0427	
444	USGS-07195500:4/12/2006:SW:S:-:-	0.0014	18	5	0.308	18	5.5	2.35	0.0521	
445	USGS-07195500:4/26/2006:SW:S:-:-	0.0014	1400	20	0.4	1400	6.7	2.33	0.0609	
446	USGS-07195500:4/26/2007:SW:S:-:-	0.0013	220	180	0.311	540	4.16	2.24	0.0395	
447	USGS-07195500:4/3/2006:SW:S:-:-	0.0012	240	24	0.234	240	5.6	2.32	0.0453	
448	USGS-07195500:4/30/2006:SW:S:-:-	0.0025			0.986		4.7	1.89	0.0878	
449	USGS-07195500:4/9/2007:SW:S:-:-	0.0008	7	1	0.119	11	4.77	2.22	0.0255	
450	USGS-07195500:5/23/2005:SW:S:-:-	0.0013			0.398	0.5	4.3	2.05	0.0606	
451	USGS-07195500:5/5/2006:SW:S:-:-	0.0048	4600	11000	2.84	4600	4.3	2.17	0.31	
452	USGS-07195500:6/12/2007:SW:S:-:-	0.0016	5400	12000	0.405	5400	4.67	2.06	0.0534	
453	USGS-07195500:6/28/2007:SW:S:-:-	0.00065	200	1400	0.175	460	5.28	2.25	0.0345	
454	USGS-07195500:6/4/2007:SW:S:-:-	0.0012	140	220	0.387	140	4.71	2.06	0.0495	
455	USGS-07195500:7/12/2005:SW:S:-:-	0.001			0.194	41	4.5	1.96	0.0448	
456	USGS-07195500:8/23/2005:SW:S:-:-	0.0009			0.308	60	6.2	2.08	0.0489	

data1 / data2 / dataones / datazeroes / pcaruns / Sheet1

Attachment C

PEA_Main_Database_Water.xls

	E	K	M	N	O	AJ	AM
1	ParamID	ValOp	Value	Rejected	ValidFlag	SampleGrp	ReasonNote
17723	Estrone	=	29	TRUE	R	HFS-16:7/11/2005:SW:S:-:BF1	Rejected data
17724	Estrone	=	98.4	TRUE	R	HFS-26:7/13/2005:SW:S:-:BF1	Rejected data
17725	Estrone	=	99.8	TRUE	R	07196500:7/11/2005:SW:S:-:-	Rejected data
17726	Estrone	=	119	TRUE	R	HFS-26:7/13/2005:SW:FS:-:BF1	Rejected data
17727	Estrone	=	122	TRUE	R	07197360:7/11/2005:SW:S:-:-	Rejected data
17728	Estrone	=	127	TRUE	R	07196090:7/12/2005:SW:S:-:-	Rejected data
17729	Estrone	=	129	TRUE	R	LK-01:7/12/2005:SW:S:1:-	Rejected data
17730	Estrone	=	137	TRUE	R	07196000:7/12/2005:SW:S:-:-	Rejected data
17731	Estrone	=	141	TRUE	R	HFS-23:7/12/2005:SW:S:-:BF1	Rejected data
17732	Fecal Coliform			FALSE		EOF01:5/14/2005:SW:S:-:-	No value reported
17733	Fecal Coliform	<	1	TRUE	R	07195500:5/23/2005:SW:S:-:-	Rejected data
17734	Fecal Coliform	<	1	TRUE	R	07196090:5/23/2005:SW:S:-:-	Rejected data
17735	Fecal Coliform	=	1	TRUE	R	HFS-04:5/25/2005:SW:S:-:1	Rejected data
17736	Fecal Coliform	<	2	TRUE	R	SPR-14:5/24/2005:SW:S:-:-	Rejected data
17737	Fecal Coliform	=	2.5	TRUE	R	SPR-36:6/2/2005:SW:S:-:-	Rejected data
17738	Fecal Coliform	=	4.6	TRUE	R	SPR-48:5/26/2005:SW:S:-:-	Rejected data
17739	Fecal Coliform	=	5	TRUE	R	SPR-04:5/25/2005:SW:S:-:-	Rejected data
17740	Fecal Coliform	=	5	TRUE	R	SPR-16:5/25/2005:SW:S:-:-	Rejected data
17741	Fecal Coliform	=	5	TRUE	R	SPR-27:5/24/2005:SW:S:-:-	Rejected data
17742	Fecal Coliform	=	7	TRUE	R	SPR-18:5/25/2005:SW:S:-:-	Rejected data
17743	Fecal Coliform	=	7.7	TRUE	R	SPR-62:6/2/2005:SW:S:-:-	Rejected data
17744	Fecal Coliform	=	8	TRUE	R	HFS-04:5/25/2005:SW:S:-:7	Rejected data
17745	Fecal Coliform	=	10	TRUE	R	07196500:5/25/2005:SW:S:-:-	Rejected data

Water (SWGW) / Water (Variables) / Water (P Protocol) / Water (Out) / USGS (N DB) / USGS (N CT)

Attachment B

Deletion Methods

The two most common deletion methods are listwise and pairwise deletion. In listwise deletion, the analysis uses complete cases only. That is, the procedure removes from computations any observation with a value missing on any variable included in the analysis.

Pairwise deletion is listwise deletion done separately for every pair of selected variables. In other words, counts, sums of squares, and sums of cross-products are computed separately for every pair of variables in the file. With pairwise deletion, you get the same correlation (covariance, etc.) for two variables containing missing data if you select them alone or with other variables containing missing data. With listwise deletion, correlations under these two circumstances may differ, depending on the pattern of missing data among the other variables in the file.

Because it makes better use of the data than listwise deletion, pairwise deletion is a popular method for computing correlations on matrices with missing data. Many regression programs include it as a standard method for computing regression estimates from a covariance or correlation matrix.

Attachment C

http://www.statsoft.com/textbook/glosp.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.statsoft.com/textbook/glosp.html

Pairwise Deletion of Missing Data vs. Mean Substitution In order to avoid loosing data due to <u>casewise deletion</u> of missing data, you can use one of two other methods. These are (1) the so-called *mean substitution* of missing data (replacing all missing data in a variable by the mean of that variable) and (2) <u>pairwise deletion</u> of missing data. These methods of handling missing data can be requested in many modules; you can also use the mean substitution method to permanently "remove" missing data from your data set. Mean substitution offers some advantages and some disadvantages as compared to <u>pairwise deletion</u>. Its main advantage is that it produces "internally consistent" sets of results ("true" correlation matrices). The main disadvantages are:

a. *Mean substitution* artificially decreases the variation of scores, and this decrease in individual variables is proportional to the number of missing data (i.e., the more missing data, the more "perfectly average scores" will be artificially added to the data set).
b. Because it substitutes missing data with artificially created "average" data points, *mean substitution* may considerably change the values of correlations.

Pairwise MD Deletion. When pairwise deletion of missing data is selected, then cases will be excluded from any calculations involving variables for which they have missing data. In the case of correlations, the correlations between each pair of variables are calculated from all cases having valid data for those two variables.

See also, <u>*Pairwise Deletion of Missing Data vs. Mean Substitution*</u> and <u>Casewise vs. pairwise deletion of missing data</u>

Exponent®

Expert Report of Brian L. Murphy, Ph.D.

Prepared for Faegre & Benson, LLP

EXHIBIT
"F"

4.4 Errors in Dr. Olsen's PCA Calculations

In conducting his PCA, Dr. Olsen used EDAnalyzer as a "shell" to run Systat statistical software. EDAnalyzer is a program developed by CDM. As indicated by Dr. Olsen, all of the calculations conducted by EDAnalyzer can be conducted outside of EDAnalyzer, and this was done to verify his results. However, following the "Steps of PCA" in Dr. Olsen's report and using his input files does not produce the results shown in his report. This is due to a fundamental error made by Dr. Olsen. PCA run SW3 is used as an example, to illustrate the nature of this error.

According to Dr. Olsen's report, the cross-tabulated data set for his PCA run SW3 was saved in a file named "Crosstab_Water_0427_SW_3.xls." This file contains the original results for the 26 variables and 573 samples identified for this analysis. The number of samples and variables matches the results provided, indicating that this file is the result of Step 8 of his "Steps of PCA" process. Appendix E of Dr. Olsen's report provides probability plots of the log-transformed values for all 26 variables. This, along with text in his report, indicates that all of the variables were additionally log-transformed as part of Step 8.

Not all samples had values for all variables, a fact that Dr. Olsen treats in an inconsistent way, as described below.

Using the data set described above as input for the PCA run, the output from Systat reproduced the coefficients reported by Dr. Olsen. This is true only when based on a correlation matrix with pairwise deletion.[6]

However, the PC scores reported by Dr. Olsen do not match the scores calculated directly by Systat, nor do they match the result of multiplying the coefficients by the standardized input data set.[7] Table 4-1 shows these three different computation results for PC1 scores for the first ten samples analyzed.

---

6 When a sample is missing a result for a variable, that sample does not contribute to the relationship between that variable and each other variable. This results in differing numbers of samples defining the relationship (i.e., correlation) between each pair of variables.

7 The standardized data set was calculated directly within Systat, because this data set was not provided explicitly within Dr. Olsen's production files. A description of this standardization was included in Dr. Olsen's report.

IN THE UNITED STATES DISTRICT COURT FOR THE

NORTHERN DISTRICT OF OKLAHOMA

W. A. DREW EDMONDSON, in his)
capacity as ATTORNEY GENERAL)
OF THE STATE OF OKLAHOMA and)
OKLAHOMA SECRETARY OF THE)
ENVIRONMENT C. MILES TOLBERT,)
in his capacity as the)
TRUSTEE FOR NATURAL RESOURCES)
FOR THE STATE OF OKLAHOMA,)
)
Plaintiff,)
)
vs.) 4:05-CV-00329-TCK-SAJ
)
TYSON FOODS, INC., et al,)
)
Defendants.)

- -

VOLUME I OF THE VIDEOTAPED

DEPOSITION OF CHARLES COWAN, PhD, produced as a witness on behalf of the Plaintiff in the above styled and numbered cause, taken on the 17th day of February, 2009, in the City of Tulsa, County of Tulsa, State of Oklahoma, before me, Lisa A. Steinmeyer, a Certified Shorthand Reporter, duly certified under and by virtue of the laws of the State of Oklahoma.

EXHIBIT
tabbies
"G"

5ffcece3-0d90-42d5-a060-5ec0426ae2da

A Your Chart 9 on Exhibit 19?
Q Yes, sir.
A It looks to me like it's six, but the other indicators that you've used are also triangles so I may be having trouble between dark green and dark blue. 04:45PM
Q And do you recall whether or not there were six reference samples that were for SW3 database?
A I'm sorry, again, I don't recall.
Q Okay. Would you now then draw a circle around those points that represent edge of field samples? 04:45PM
A I'm sorry, again on Chart 9?
Q Yes. Thank you.
A That's okay.
Q They're the blue diamonds I believe, sir. 04:45PM
A Yes. Well, actually I should have said this after the last grouping, too, but what you're -- let me make sure that I understand what you're asking me. You're asking me to create a group that contains -- 04:46PM
Q Edge of field.
A -- most of them or all of them.
Q All of them that you can see.
A Okay.
Q Now, when you've done that, sir, can you see 04:46PM

there's a distinction where the samples that I've represented to you are wastewater treatment plant are separated from those that are edge of field samples and separated from those that are reference samples? 04:46PM

MR. TODD: Object to form.

MS. COLLINS: Object to form.

A Well, I see that there are three non-overlapping groups that were formed by my circles. Is that your question? 04:47PM

Q Yes, sir.

A Yes.

Q And what did you understand a reference sample to be in Dr. Olsen's database?

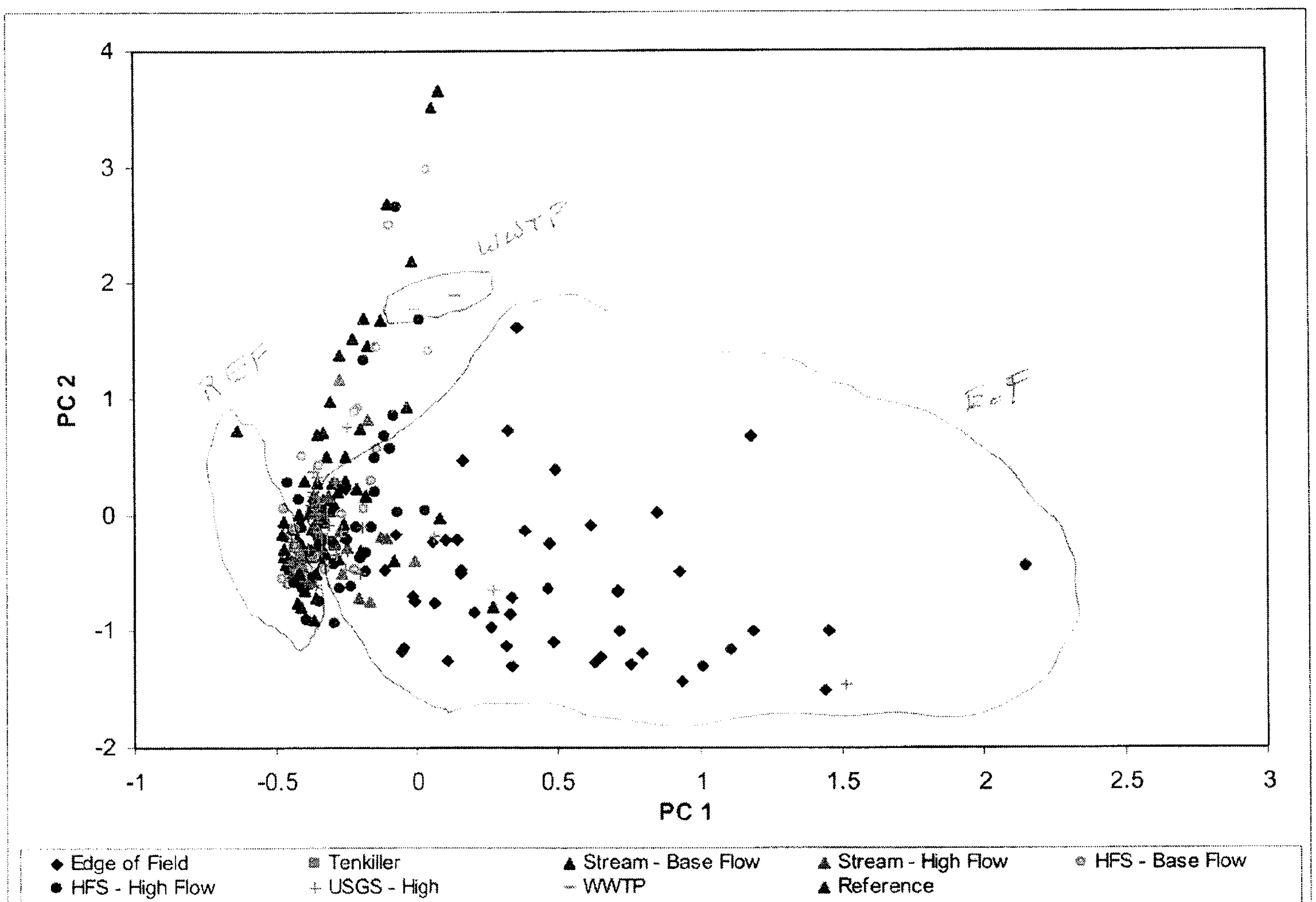
A As I understood it, it was a sample taken 04:47PM where it was known or alleged that there was -- there were no poultry farms nearby.

Q Okay. Was it also important that there were no other sources of contamination, such as wastewater treatment plant, where the reference 04:47PM sample was taken?

A I believe that that's correct.

Q Okay, and turning now to the wastewater treatment plant samples, what was your understanding of the purpose of those samples? 04:47PM

Chart 9 – Colored Coded



COWAN
DEPOSITION EX# 19

Chart 10 – Colored Coded